

James M. Hyman  
 Thomas A. Manteuffel

Los Alamos National Laboratory  
 Los Alamos, NM 87545

# ABSTRACT

We are interested in solving the sparse linear systems,  $Av = b$ , that arise from finite difference or finite element approximations to partial differential equations. Many iterative methods require solving an easier approximate equation,  $Pv = b$ , on each iteration. This is often called preconditioning or operator splitting [1,2,4,6-9,12-14]. The methods we consider factor  $A$  approximately into the product of an upper and lower triangular matrix  $P \approx LU = A$ . These methods are called incomplete LU factorization methods and their convergence rate depends on how well  $P$  approximates  $A$ . We describe some new algorithms to generate accurate LU decompositions based on the continuity of the solution  $v$ .

## I. INTRODUCTION

In this paper we consider iterative numerical methods for the solution of sparse linear systems of the form,

$$Av = b, \quad (1)$$

that arise in the finite difference or finite element approximation of a partial differential equation (PDEs). In these equations the vector  $v$  of length  $N$  approximates the smooth solution to a discretized PDE defined on an underlying  $n \times n$  spatial mesh. The smoothness of the solution will be exploited to define high order approximate factorizations of  $A$ . For simplicity we will analyze these problems for two-dimensional equations where  $N = n^2$  but the methods generalize easily to higher dimensions.

If the matrix  $P$  approximates  $A$ , then nonstationary iterative methods of the form [15]

$$Pv^{(i+1)} = Pv^{(i)} + \zeta_i(b - Av^{(i)}), \quad (2)$$

can be used to solve (1.1). The matrix  $P$ , called the preconditioning matrix, is generated by approximating  $A$  by an incomplete LU decomposition [2,6,8,9,12-14]. That is,

$$A = LDU - R = P - R, \quad (3)$$

where  $L$  is lower triangular,  $D$  is diagonal,  $U$  is an upper triangular matrix and  $R$  is the residual error matrix. The matrices  $L$  and  $U$  have unit diagonals. These matrices are chosen so that the residual matrix  $R$  is small and the approximating system  $Pv = b$  is easier to solve than Eq. (1). The nonzero structure of  $L + U$  is chosen to be similar to (if not equal to) the nonzero structure of  $A$ .

The number of iterations the algorithm takes to converge to a specified error tolerance is related to the condition number of  $P^{-1}A \equiv \kappa(P^{-1}A) = \kappa(I - P^{-1}R)$ . (The condition number is the ratio of the largest to smallest singular value of a matrix.) Note that the smaller the residual matrix is the smaller the condition number will be and the faster the method will converge.

Often the iteration can be accelerated using a conjugate gradient [1,3-5,7,8,12], Chebychev [10,11], or other polynomial acceleration algorithm. When conjugate gradient acceleration or Chebychev acceleration is used the convergence rate is proportional to the square root of the condition number.

We will give a non-rigorous argument on how to estimate the condition number  $\kappa(P^{-1}A)$ . We assume  $P$  is close enough to  $A$  so that if  $\|A^{-1}\| = O(h^{-\alpha})$  then  $\|P^{-1}\| = O(h^{-\alpha})$ , where  $\|\cdot\|$  is the operator norm induced by the 2-norm and  $h = 1/n$  is the mesh spacing. (For Laplace's equation, we have  $\alpha = 2$ .) When  $\|R\| = \|A - P\| = O(h^\beta)$  then the condition number of  $P^{-1}A$  can be approximated by

$$\begin{aligned} \kappa(P^{-1}A) &= \|P^{-1}A\| \|A^{-1}P\| \\ &= \|I + P^{-1}R\| \|I - A^{-1}R\| \\ &\approx [1 + O(h^{-\alpha}) O(h^\beta)][1 + O(h^{-\alpha}) O(h^\beta)] \\ &= 1 + O(h^{\beta-\alpha}) \\ &= 1 + O(N^{(\alpha-\beta)/2}) \end{aligned}$$

The condition number of  $A$ ,  $\kappa(A) = O(h^{-\alpha})$ , is determined by the problem being solved but the condition number of the residual matrix can be reduced by making  $P$  as close to  $A$  as possible. That is, if  $\|R\| = O(h^\beta)$ , then we wish to make  $\beta$  as large as possible. In particular, if  $\beta > \alpha$  then the method will converge faster for larger problems than smaller ones.

A standard method for forming  $L$  and  $U$  is to perform a complete decomposition and discard fill-in [12] or to add the discarded fill-in back to the diagonal [8,9]. We wish to abandon this approach in favor of treating the elements of  $L$ ,  $D$ , and  $U$  as unknowns and choosing them so that  $R$  has certain desirable properties.

Since  $A$  is assumed to be the discretization of a partial differential operator we know that the eigenvectors (singular vectors) associated with the smallest eigenvalues (singular values) will be relatively smooth. The components of the error in the direction of these eigenvectors will be the most difficult to resolve using iterative methods. For this reason, we would like to construct  $P$  to approximate  $A$  closely on the discrete analogues of smooth functions. In other words, we will choose  $P$  so that  $R$  is as small as possible on smooth functions.

To illustrate this approach consider the discretization of the diffusion convection equation

$$-\frac{\partial}{\partial x}(\alpha_1 \frac{\partial v}{\partial x}) - \frac{\partial}{\partial y}(\alpha_2 \frac{\partial v}{\partial y}) + \beta_1 \frac{\partial v}{\partial x} + \beta_2 \frac{\partial v}{\partial y} + \gamma v = f \quad (4)$$

defined on the region  $[0,1] \times [0,1]$ .



The five-point star second-order approximation to Eq. (4) often leads to a finite difference approximation that can be written

$$\begin{aligned} & a_{-1,0}^{k-1} v_{i-1,j} + a_{0,0}^k v_{i,j} + a_{1,0}^k v_{i+1,j} \\ & + a_{0,1}^k v_{i,j+1} + a_{0,-1}^{k-n} v_{i,j-1} = b_{i,j} \end{aligned} \quad (5)$$

at each point in the mesh. There  $v_{i,j}$  is the solution at  $(x_i, y_j)$  and is the  $k$ -th  $[k = i + (j - 1)n]$  element in the one dimensional  $v$  array.

The matrix  $A$  is shown in Fig. 1 and will be symmetric ( $A = A^T$ ) if  $a_{1,0}^k = a_{-1,0}^{k+1}$  and  $a_{0,1}^k = a_{0,-1}^{k+n}$ .

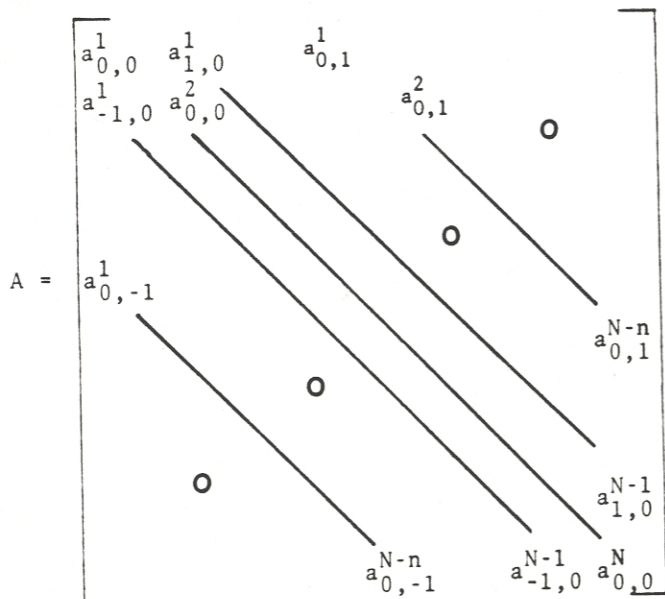


Figure 1. The matrix  $A$  for the five point operator.

The stencil of a finite difference approximation can be an excellent tool to understand how the underlying mesh, the matrix  $A$ , and its approximation  $P$  are related. Using notation from Dupont, Kendall and Rachford [6] a five point approximation to (4) leads to a matrix  $A$  described by the stencil in Fig. 2.

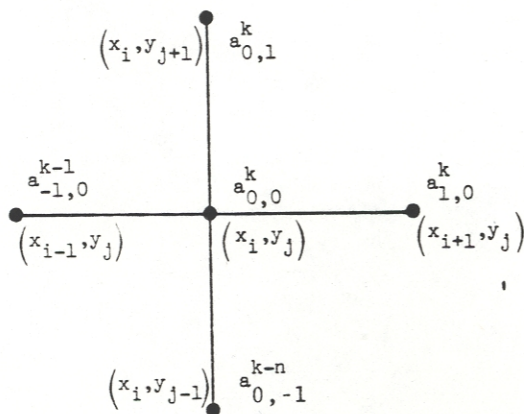


Figure 2. The stencil for the matrix  $A$ .

The coefficients at the nodes (mesh points) are the elements in the  $k$ -th,  $[k = i + (j - 1)n]$ , row of  $A$ .

In the simplest incomplete LU decomposition of  $A$ , the matrix  $L + U$  has the same nonzero structure as  $A$ . We shall call this factorization the  $ILU(0)$  method since it has no extra nonzero elements.

#### The $ILU(0)$ Method

In the  $ILU(0)$  method the lower triangular matrix  $L$  shown in Fig. 3.a, has the stencil shown in Fig. 3.b.

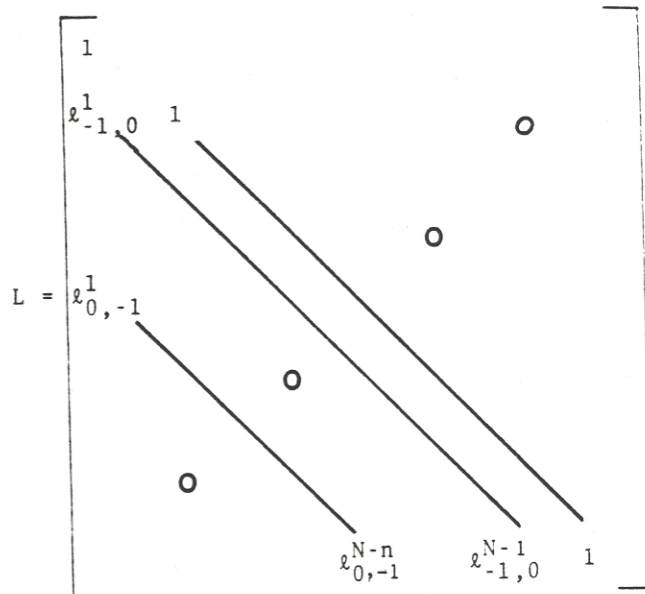


Figure 3.a. The matrix  $L$  for the  $ILU(0)$  method.

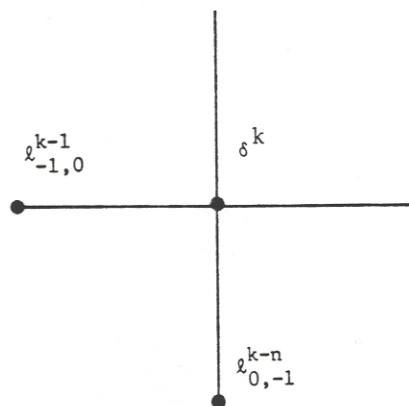


Figure 3.b. The stencil for  $L$ .

The upper triangular matrix  $U$ , shown in Fig. 4.a, has the stencil in Fig. 4.b. The diagonal matrix  $D$  has diagonal elements  $\delta^k$ . The stencil for the preconditioning matrix  $P = LDU$  is shown in Fig. 5.



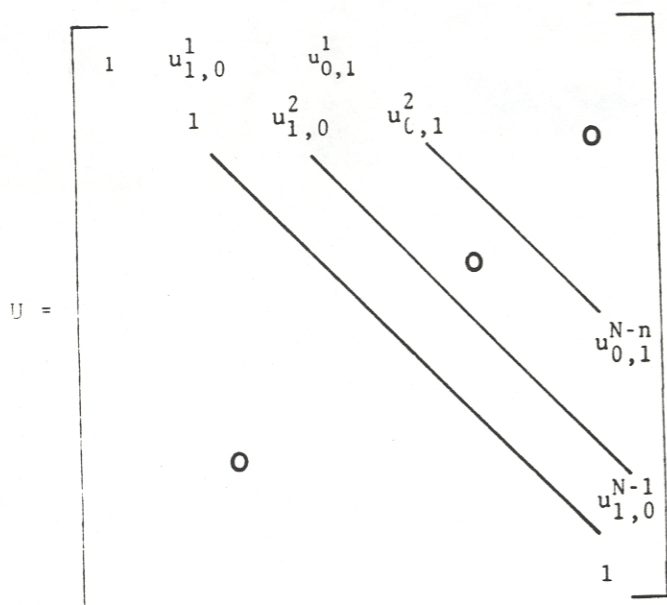


Figure 4a. The matrix U for the ILU(0) method.

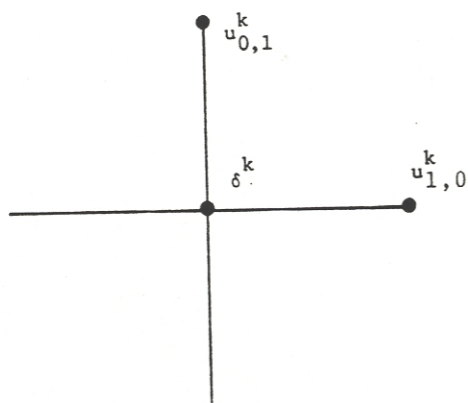


Figure 4b. The stencil for U.

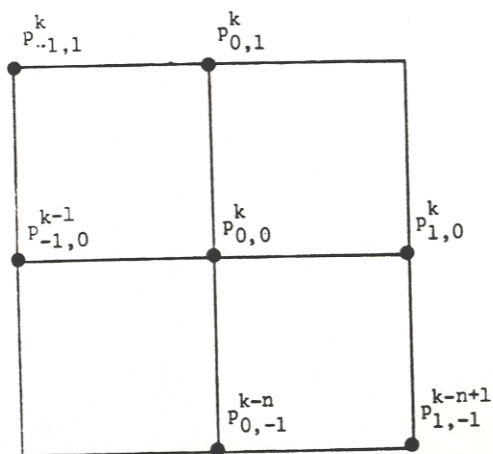


Figure 5. The stencil for the preconditioning matrix  $P = LDU$ .

This will also be the stencil for the residual matrix  $R = P-A$ . Each entry in the  $k^{\text{th}}$  row of  $R$  will

be a linear function of the five unknowns  $\delta^k, \ell^{k-1}, \ell^{k-n}, u_{1,0}^k, u_{0,1}^k$ . These terms are also functions of other values of  $\delta, \ell$ , and  $u$  which we shall assume to be chosen already.

Since there are seven stencil points, it will not be possible in general to make the  $k^{\text{th}}$  row of  $R$  identically zero. We may, however, choose the parameters so as to make the stencil zero on the points associated with the stencil of  $A$ . This corresponds to the methods in [10]. If we choose the unknowns so that the product of  $R$  times a constant vector is zero we have lost only one degree of freedom. One choice of the remaining four corresponds to the method described in [8]. Suppose we use two more degrees of freedom to make the product  $Rv = 0$  when  $v$  corresponds to a function that is constant in  $x$  and linear in  $y$ , or when  $v$  corresponds to a function that is constant in  $y$  and linear in  $x$ ; then, there are two degrees of freedom left. If  $v$  corresponds to a function with two continuous derivatives each term of the product  $Rv$  will be  $O(h^2)$ .

The remaining two degrees of freedom are insufficient to make the product zero on the quadratics and bilinear functions, consequently  $O(h^2)$  is the best we may achieve. However, the two degrees of freedom may be used to reduce the size of the elements in the  $k^{\text{th}}$  row of  $R$ . If we choose to minimize the sum of the squares the unknowns may be found by solving a  $5 \times 7$  constrained least squares problem. Another use of these degrees of freedom might be to make  $L^T - U$  as small as possible and thus  $P$  more nearly symmetric.

The entire procedure involves passing sequentially through the grid points and choosing the five unknowns so that  $(Rv)_k = O(h^2)$ . At each grid-point, the equations involve elements of  $L, D$ , and  $U$  that were chosen at previous grid-points, and the five elements mentioned above. This preconditioning has been shown to be superior to incomplete Cholesky [12] or modified incomplete Cholesky [8] on test problems. However, what we really seek is a factorization for which  $(Rv)_k = O(h^3)$ . If the elements of  $R$  remain bounded as  $h$  is reduced to zero, this would yield  $\|Rv\| = O(h^2)$  whenever  $u$  corresponds to a smooth function. The condition of  $P^{-1}A$  would be independent of  $h$ . For this we need a larger stencil.

#### The Compact ILU(1) Method

A more accurate incomplete factorization of  $A$  can be formed by allowing  $L$  and  $U$  to have nonzero elements in the positions other than those where  $A$  is nonzero. If both  $L$  and  $U$  have one extra nonzero element then the method is called an ILU(1) method. There are several choices as to where this extra element can be added. To reduce the extrapolation error we chose to add this extra element such that the stencil for  $P = LDU$  will be as close to  $A$  as possible. This choice, called the compact ILU(1) method, has an  $L$  and  $U$  with the nonzero structure shown in Fig. 6. (The factorizations described in [8] and [12] add  $\ell_{+1,1}^{k-n+1}, u_{-1,-1}^k$  to the stencil of  $L$  and  $U$ .)



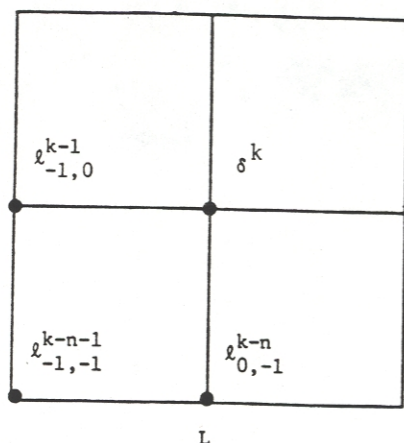


Figure 6a. The nonzero structure for L for the compact ILU(1) method.

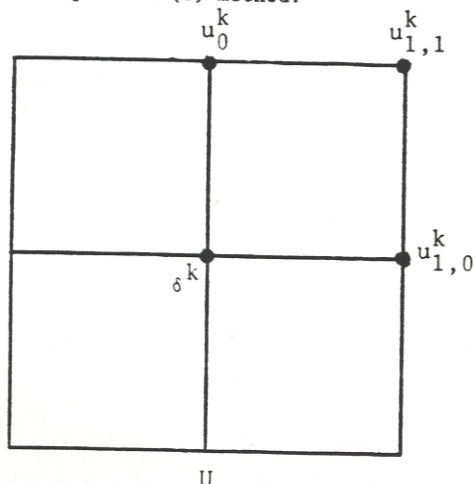


Figure 6b. The nonzero structure for U for the compact ILU(1) method.

Their product  $P = LDU$  has the nonzero structure shown in Fig. 7, as does the residual matrix R.

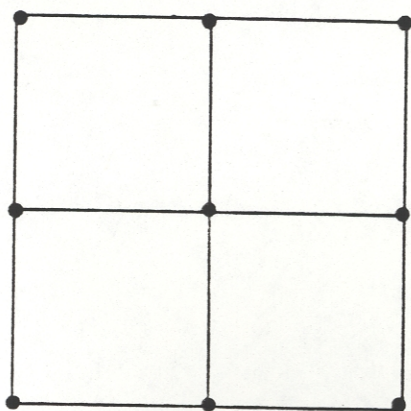


Figure 7. The nonzero structure for  $P = LDU$  for the compact ILU(1) method.

With this stencil we have seven unknowns and nine stencil points. Again R cannot be made identically zero. Suppose  $v$  corresponds to a quadratic polynomial; that is,  $v(x,y) = p_0 + p_1x + p_2y + p_3x^2 + p_4xy + p_5y^2$ . Let us choose the elements of R so that  $Rv = 0$  for any such polynomial. This requires 6 degrees of freedom. The remaining degree of freedom may be used to minimize the sums of the

squares of the elements of R or to promote the symmetry of P. These conditions yield a  $9 \times 7$  constrained least squares problem. We may in this fashion pass sequentially through the mesh choosing L, D, and U in such a way that if  $v$  corresponds to a function with three continuous derivatives then  $(Rv)_k = O(h^3)$ .

Unfortunately, the size of the elements of R grows as  $h \rightarrow 0$  with the net result that  $\|Rv\| = O(h^2)$ . If instead we use only three degrees of freedom to make  $(Rv)_k = O(h^2)$  and the remaining four to reduce the size of the elements of R, a preconditioning results that is again superior to IC and MIC. The other four degrees of freedom may be used to promote symmetry and, in fact, P can be made symmetric except near two of the four boundaries.

#### The ILU(2) and ILU(3) Methods

Let us consider even larger stencils in the hope that one may be found for which  $\|Rv\| = O(h^3)$  for sufficiently smooth  $v$ . Clearly, such a stencil exists because exact decomposition yields  $Rv \equiv 0$ . In the ILU(2) method we add one more point to the stencil of L and U in the positions shown in Fig. 8.

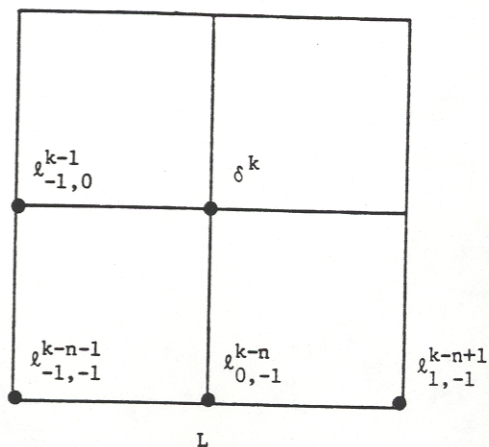


Figure 8a. The stencil for L for the nine point ILU(2) method.

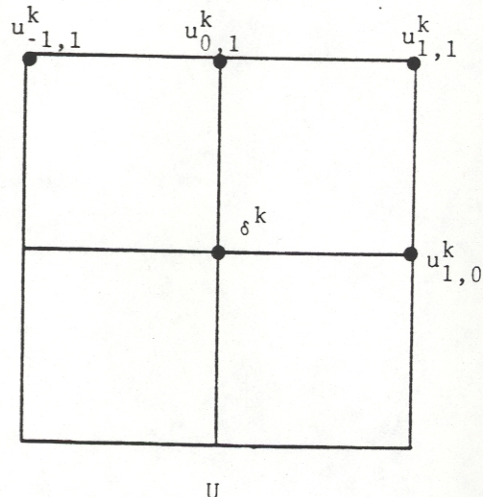


Figure 8b. The stencil for U for the nine point ILU(2) method.

The stencil for  $P = LDU$  is shown in Fig. 9.



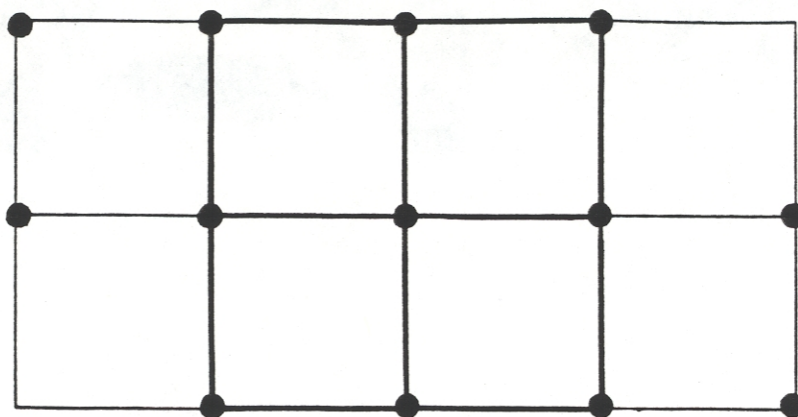


Figure 9. The stencil for  $P = LDU$  for the nine point ILU(2) method.

Using this same strategy we can add three extra points to the basic ILU(0) stencils. In this ILU(3) method there are 11 unknowns and 15 stencil points. Recall that only six degrees of freedom are required in order to make  $(Rv)_k = O(h^3)$  on smooth functions. The remaining degrees of freedom can be used to control the size of the elements of  $R$ .

#### Summary

In this paper we have presented a heuristic motivation for choosing the elements of a family of incomplete factorizations. Members of this family have been shown to be numerically superior to the IC(12) and MIC(8) factorizations. However, the work required to form the factorizations is greater and not easily vectorized. If a factorization could be found for which  $\|Rv\| = O(h^3)$  on smooth functions then the extra work could be justified. Numerical results for the ILU(2) and ILU(3) methods are incomplete but show great promise.

#### REFERENCES

1. O. Axelsson and I. Gustafsson. "A modified upwind scheme for convective transport equations and the use of a conjugate gradient method for the solution of non-symmetric systems of equations." Computer Sciences 77.12R, Chalmers University of Technology, Göteborg, Sweden (1977).
2. R. Beauwens and L. Quenon. "Existence criteria for partial matrix factorizations in iterative methods." SIAM Journal on Numerical Analysis 13:615-643, 1976.
3. R. Chandra, S. C. Eisenstat, and M. H. Schultz. "The modified conjugate residual method for partial differential equations." Proceedings of the Second IMACS(AICA) International Symposium on Computer Methods for Partial Differential Equations (Ed. R. Vichnevetsky), Lehigh University, Bethlehem, Pennsylvania, 1977, pp. 13-19.
4. P. Concus, G. H. Golub, and D. P. O'Leary. "Numerical Solution of Nonlinear Elliptic Partial Differential Equations by a Generalized Conjugate Gradient Method," Computing 19, 321-339 (1978).
5. P. Concus and G. H. Golub. "A generalized conjugate gradient method for nonsymmetric systems of linear equations." Report STAN-SC-76-535, Department of Computer Science, Stanford University, 1976.
6. T. Dupont, R. P. Kendall, and H. H. Rachford. "An approximate factorization procedure for solving self-adjoint elliptic difference equations." SIAM Journal on Numerical Analysis 5:559-573, 1968.
7. S. C. Eisenstat. "Efficient Implementation of a Class of Preconditioned Conjugate Methods." Yale Computer Science Research Report No. 185, August 1980.
8. I. Gustafsson. "A class of first order factorization methods." BIT 18:142-156, 1978.
9. J. M. Hyman, "High-Order Incomplete LU Factorizations of Sparse Matrices," Los Alamos National Laboratory Report LA-UR-80-3701 (1980).
10. T. A. Manteuffel. "The Tchebychev Iteration for Nonsymmetric Linear Systems." Numerische Mathematik, Vol. 28,307-327 (1977).
11. T. A. Manteuffel. "Adaptive Procedure for Estimating Parameters for the Nonsymmetric Tchebychev Iteration." Numerische Mathematik, Vol. 31, 183-208 (1978).
12. J. A. Meijerink and H. A. van der Vorst. "An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix." Mathematics of Computation 31: 148-162, 1977.
13. P. Saylor. "Second order strongly implicit symmetric factorization methods for the solution of elliptic difference equations. SIAM J. Numer. Anal. 11:894-908, 1974.
14. H. L. Stone. "Iterative solution of implicit approximation of multidimensional partial differential equations." SIAM Journal on Numerical Analysis 5:530-558, 1968.
15. D. M. Young. Iterative solution of large linear systems. Academic Press, 1971.